

AT THE EDGE OF UNCERTAINTY IN LANGUAGE

Inger Bierschenk
Psychological Laboratory, University of Copenhagen
and
Copenhagen Competence Research Center

Abstract

This article presents the prerequisites and implications of a new method called Perspective Text Analysis. The discussion concerns some basic differences in social science and linguistic approaches to language modelling and its consequences. The concepts of organisation and structure are contrasted and it is shown and explained the reason why these concepts are mixed up in models of language and texts. In the model presented, text is conceptualised as an exact depiction of the continuous and irreversible flow of information coming into and going out of an agent's sight. The agent as biological phenomenon is the fundamental property of the model, which frees the researcher from dealing with artificial concepts, constructivism, and conventionalism. A biologically rooted mechanism for text production is presented, which operates at the cutting edge of organisation and structure.

Text and context

A great majority of scientific text analyses is initiated on the basis of the researcher's assumptions of what should be expressed in a certain text within a particular area. In a broad social science perspective there are as many methods of text analyses as there are projects or fields of research. Methods with, for example social, psychological, economic and political constraints are continuously presented and discussed at nearly all national and international symposia. This is due to the fact that the researcher or research group develops their own theories of the "structure" of a field. The work is aimed at affirming or, in the single cases, rejecting this structure by the method used. A common name for a lot of such methods is "content analysis".

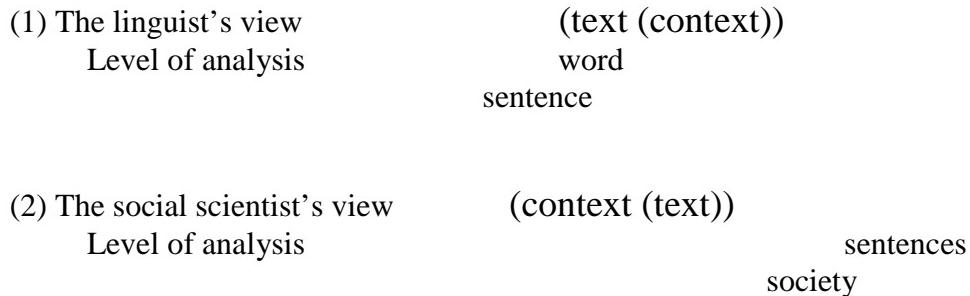
Each content analysis implies some kind of pre-defined classification on the basis of which a categorisation procedure takes place. It ends up with pieces of text, void of conceptual connection once the contextual links have been cut off. Everyone who has gone through this procedure is aware of its weak reliability. The entire coding process is highly time consuming, at least by larger text materials, and when performed it is hard to apply anew, among other things because the analyser during processing quite simply is not able to keep in mind the categorisations made at an earlier stage. It is even more difficult to perfectly agree with another analyser. Consequently, to this methodology is connected a stock of reliability tests, which make the result a question of probabilities, in the statistical sense.

Since the early 70's and for at least a decade I worked with great amounts of interview materials (4000 pages), consisting of answers to unbound questions (Bierschenk & Bierschenk, 1976). I know what it means to develop coding rules and write a manual and have all the rules and decisions tested for coder agreement (I. Bierschenk, 1977). With this

experiential background, I would like to state that the final result of all this is the personal feeling of having learnt a lot about text production and layout and about different processes involved in the various steps of analysis. But it became more and more difficult to see how this layout or “textual face” could be capable of reflecting something internal, that is, “structure”.

It is a known fact that linguists and social scientists approach texts very differently, at least traditionally seen. To a linguist, the text includes its context, while to a social scientist, the context may imply the wider frame of interpretation, that is, the text is included in the context (Fig. 1).

Figure 1. The text concept in linguistics and social science



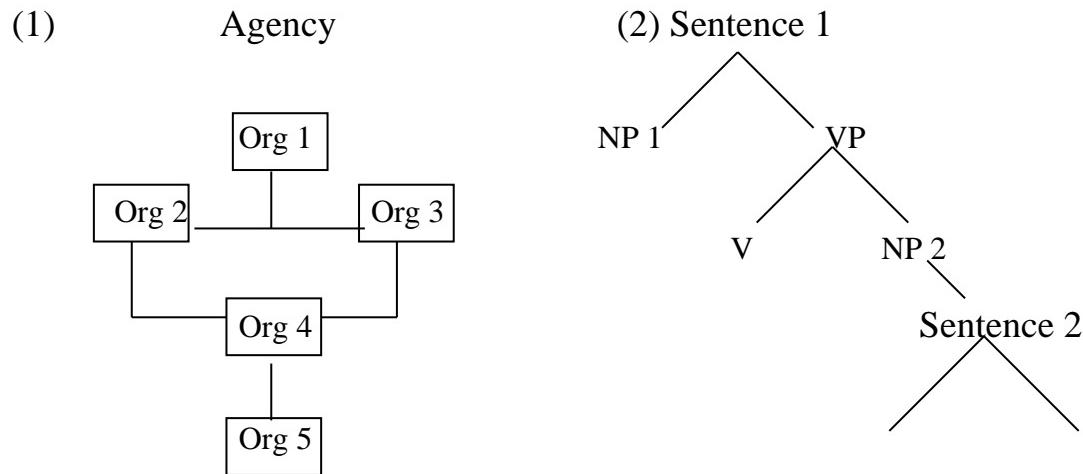
Somewhere within those borderlines the linguist and the social scientist meet under a joint concept, namely “structure”, which by both is conceived of as an internal quality of text, but which they try to capture by looking at its face, that is, the “organisation”. As a consequence, the same basic idea lies behind “content analysis” and “semantics”.

Text as organisation

Typical of “organisation” is the definition of units and their interrelations, often described in terms of levels. Identification and analysis is carried out by means of a rule system departing from a base or kernel and associating the units into a certain order so that the procedure of reaching complete overview will be as efficient as connectivity can be. Depending on which complex relations the organisation stands for, the path from top to base may be more or less linear. There are parallel orderings and intermediate levels, more or less important nodes to pass. Sometimes decisions have to be made, for example, because some units may have the same function but different labels or vice versa. Other units are just slot fillers in the building and function as barriers. Common to organisations of various kinds is that they are visual and manifest, at least they have visual representatives making hidden parts easily reachable, for example by reference and inference.

One consequence of the social science view on text as organisation is to regard the text as part of a kind of context formed by a collocation of organisations, which we might call an *agency*. Seen in this way one text and its agent is not an autonomous whole but is linked, let be in arbitrary order, to an agency of texts, whose common denominator the analysis shall detect. Seen in a narrower linguistic context the

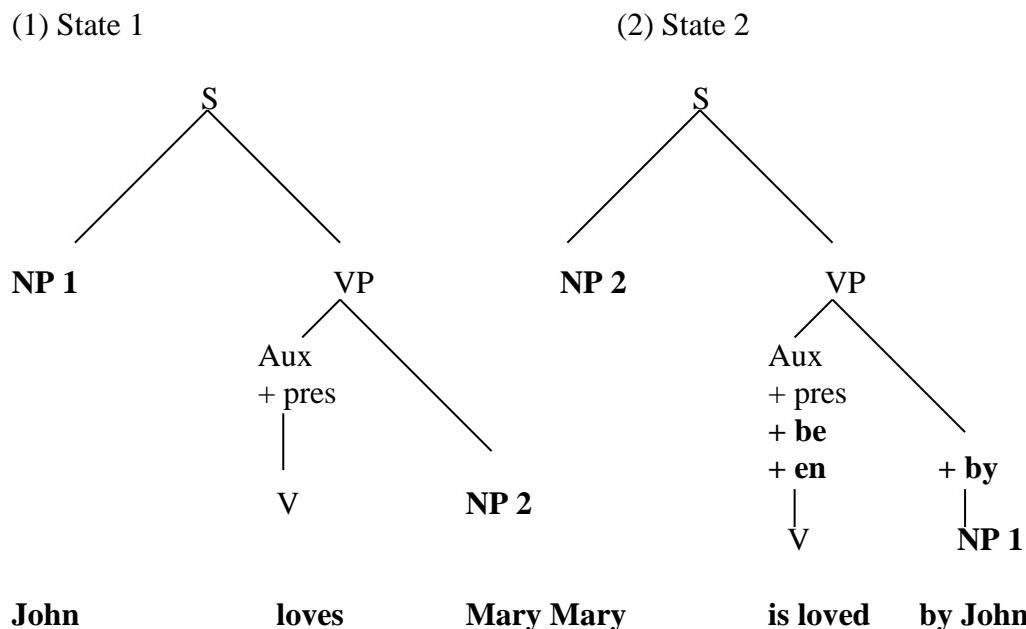
Figure 2. Text seen as organisation



organisation is a grammatical sentence and its parts. Some parts are main components while some are subordinated.

I am convinced that most of the social scientists use the word “structure” when they mean order and relations, just like linguists have done since Chomsky (1957) proposed his description of a sentence as “syntactic structures”. How then can we explain the mixing up of organisation and structure? At least in traditional linguistics a kernel is assumed from out of which expansions and transformations can be observed. *Transformation* is a structural concept, but in sentence analysis it is used in its classical mechanical sense, that is, for positional change as for example in the passive construction. Figure 3 shows an analysis of the prototypical textbook example of how the active “John loves Mary” is “transformed” into the passive “Mary is loved by John”.

Figure 3. The passive construction as positional change



As can be seen, there is a change of level in the NP's position at state 2, the passive verb form adds a couple of features, and a preposition marks the semantic identity of NP 1 in spite of its positional change.

The idea to prove structural development via syntactic transformation caused a flow of psycho-linguistic and psychological tests in the 60's and 70's. One hypothesis tested was whether the degree of transformation (i.e., depth) on the kernel sentence could be associated with degree of difficulty in understanding, measured for example as processing time (RTM) in reading the sentence. Broadbent's (1977) comprehensive review shows that this is not the case.

To conclude - organisational change cannot guarantee structural control whether the steering component be text or not.

Text as pre-defined knowledge

A text analysis, whose fundamental governing concept is structure, has to keep track of the functions and channels through which the steering component (the source) is operating. In several social science models we find the concept of an *actor*, sometimes standing for the status of an individual or organisation. However, typical of an actor is only representation, because function and position are not separated. It follows that "the form", that is, the *role*, is central to such a model.

The same idea is to be found in the linguistic model. A so-called grammatical role builds on positional thinking and that the various role representatives are similar in the sense of classification: They must be of the same type to fit into the role. Quite natural, the analysis of a "Gestalt" is based on pre-defined knowledge of the scenarios, the frames for the roles (I. Bierschenk, 1984). I would like to show, by a simplified example, the way in which this actor - role model has effected the view of language and text analysis (Fig 4.) The different units of the sentence (taken from an authentic interview) will have specified grammatical roles. For the Scandinavian languages, Diderichsen (1962) worked out a so called "position scheme" for description of types of sentences and clauses.

Figure 4. Grammatical role analysis

Text Analysis 1	Why q-word	shall finite verb	I pronoun	help infin. verb	the def. art.	community noun
2			Pred 1	Subject	Pred 2	Object

The columns represent positions, which can be grouped together into *fields*, more or less closely related to one another. Thus the relation between Pred 1 and Subject, called the *nexus*, is the most tight one acting like a kernel around which other fields are positioned. The infinite verb has another role than the finite and as a second predicate it belongs to a "content field".

The field approach connects to European structuralism in psychology where concepts like "dependency" and "developmental constraints" are central. Applied to language, especially in analysing human verbal behaviour, the translation of structure concepts has so far not been successful. One such translation problem concerns *schema*. It means the genetically rooted device for information pickup and structuring, physically as well as mentally, when used as a

scientific concept. In linguistics as well as in social science, it is, however, used in the sense of scheme or plan and thereby it becomes synonymous with frame, which is an organisational concept. (For the schema concept in operation I refer to the next section.)

The semantic elaboration of the “syntactic structures” approach takes into account the pre-defined grammatical roles of nouns and the prediction it gives for automatic analysis. One such problem for computational linguistics has been to algorithmically decide whether a certain noun is the subject of a clause or not. Behind the problem lies the fact that a logical proposition ($v(NP1, NP2)$) does not differentiate between the first and second noun since it is non-directional. Fillmore (1968) worked out a hierarchical system in which the semantic case (Agent, Instrument, Object, etc.) qualifies the order in which nouns can take the role of (act as) subject, i.e., first noun. The case, in its turn, is dependent on so called “case frames” for verbs, which implies a listing of pre-defined knowledge of what kind of action the sentence denotes and what case roles are possible. Clearly, the semantic network that such knowledge requires seems almost infinite.

There are also great doubts as to the logic underlying the definition of the semantic case features. The agent, for example, may only be a human or an animal, according to the common knowledge model. With this frame of reference it is easy to realise that the analysis could only deal with single sentences of a very logical and artificial kind and not with texts. The Chomskyan model owes very much to ancient language philosophy. It is evident how for example Panini's grammar has had an influence on Fillmore's cases. However, this grammar, although clear and systematic, is known as being hierarchical and static (Singh, 1979). So, none of the traditional approaches, neither modern nor ancient, would be capable of handling a non-expected action. This would namely require a model designed on the basis of principles establishing “flow stability in complex unknown surroundings” (Elstrup Rasmussen, 1997).

Text as structure

I think that anyone who has some experience of an organisation knows that the informal ways are much more important than every formal step or stair. In places involving people there is also a dynamics whose operations are both rules breaking and unexpected. Connected to a dynamic functioning is above all open channels, possibility for functional change and sensitivity to “the spirit” (“Geist”). These “streams” in the system are necessary for development to come about. However, to gain good results from the streams, it seems as if the best would be to have at least some positions filled by certain key functions, which do not change. In this way the organisation's pathways and goals will be clarified, that is, position and function support each other.

So, the structure is only partly visible and often just by chance through organisational keys. But since these are stable, we can get at the pulse, the movement, and the power without which the organisation is nothing more than an empty shell. In a structure model applied onto language and text, a basic assumption should be that the components depict the process of *perspectivation*, which implies the discernment of a source and an operator, separated but co-operating in building up a perspective. A perspective depicted in text should be a mentally ordered visual field, which has implications on the steering and control functions of textual movement, or *flow*. A text analysis whose fundamental property is to detect and depict the perspective hidden in textual flows actually exists.

Perspective Text Analysis (PTA) is the name of a method, which began developing at the end of the 1970's. It is the result of collaboration between psychology and linguistics at Lund University. PTA together with the PC system named PERTEX and a method for topological representation makes up the Scanator. For a general presentation, I refer to e.g. Bierschenk, I. (1984, 1992); Bierschenk, B. (1991, 1993); Bierschenk, Bierschenk, & Helmersson, (1996), and Helmersson (1992).

To make clear in what way a perspective is present in a text, I have to replace the actor by an *agent*. Hopefully I will be able to show that the structural concept of agent operates very differently compared to its semantic counterpart. Agent stands for an individual, autonomous way of steering a process. Depending on what the process is about or where in the flow the steering mechanism is visible, its source manifests itself variably and in the most appropriate shape. Thus a sub component or part may stand for a whole, which means that the whole in all its parts is not known but develops during processing. An important consequence of this model is that the agent function is bound to the first position in a functional schema. A schema differs from a frame insofar as any textual information may take the function of agent. In this way it is possible to control the co-operation between the visible and non-visible, or unknown.

The concept of agent is probably known to a natural scientist as some cause or potential energy, external to the object of study. Used in this sense the agent of a text would be something or someone that causes the text to be written but may not be part of the text. The agent of the present model differs from this conception in one fundamental respect: It is part of the model. PTA regards text building as biologically founded behaviour and therefore it is impossible to disregard the agent, that is, the human being that is the very root of this behaviour. Because the root is internal to text it has been incorporated into the model too. This conception, which connects to the schematism of Kant, was first presented in B. Bierschenk (1984, p 11) and has been further discussed in relation to the consequences for text processing in I. Bierschenk (1989, p 25).

The most striking implications of the

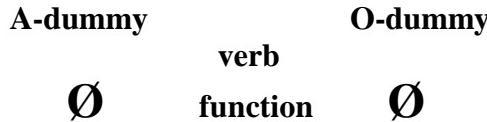
(A a (A a O))

model for the study of human language are the following:

1. Language is *biological*. As any other *natural* phenomenon it is subject of evolution. It would thus be against natural law to treat language as a construction, as traded convention. Anyone who has studied text production in more detail can confirm that this complicated phenomenon would never have been invented. So why study it by means of artificially defined models?
2. We can study language as any human body, and the result concerning one or a few can be generalised to all the others.
3. The incorporation of the agent into the model makes the analysis free from external reference. The analysis becomes totally objective.

A main difference compared to the traditional linguistic sentence (sometimes clause) analysis is expressed by what I have termed “the functional clause” (I. Bierschenk, 1992, p 7). Figure 5 presents the functional clause, which at the same time represents the schema in operation.

Figure 5. The functional clause of perspective text analysis



Applied on a text, the functional approach means that there is an agent present before the verb (read from the top) and an object after it. The presence, however, does not mean that variables are textually visible at their functional place. But with every verb, here treated as a constant, there is a key to a functional clause. To take this presence into account means to define a borderline where the certain and the uncertain meet. Through the functionally bound positions PTA detects whether they are filled with variables or not. When empty, this means that the agent hides itself or its objective, or both. The task of the algorithmic analysis therefore is to pick up the hidden information and insert it at its functional place. This principle of text supplementation is shown in Figure 6.

Figure 6. The text supplementation principle

He wanted \emptyset to \emptyset see the book	Agent v Object Agent v Object	Supplementation downwards \emptyset He + the book He Supplementation upwards
--	--	--

The text producer directs the text forwards, which means that information flows irreversibly. There is no such concept as inverted word order. Thus what would be a semantic - logical subject or agent may functionally be the object in the perspective of the irreversible flow. To show the difference in approach, I turn to the example of Figure 4 and give in the next Figure the functional analysis.

Figure 7. Functional analysis

Text	Why \emptyset shall I \emptyset help the community X verb O X verb O
Analysis	

At the same time as the textual information is not visible this circumstance opens up a peephole and channel to a space. This elicits a flowing of information through the text, since in that case the unknown agent (X) performs its government from another place. Through the hole, X can be detected. The bound position becomes a window through which a text producer chooses to show up or not, or through which a “Geist” may stream, impossible to grip in advance, but fully developed and conceivable when it has reached the end. The variable agents, also called textual agents, have the function in common that they regulate the streams descending from the X-agent. Any other similarities between them are not possible to state a priori. Thus it is the functional use of the agent that paves the way for perspective information. The relationship between textual agents, text producer and some spirit or

super-ordinate idea does not build on classes and categories but on functional invariants, in the Kantian sense.

I hope to have been able to show some distinct difference between the concepts of organisation and structure. Even though the models within social science and linguistics put different perspectives on language and text I see a way of joining the two areas by means of the concept of agent. Agent and agency lie in the same visual field, such that agency is a generalisation of agent. An analysis with agency as steering would then result in wider concepts than an analysis steered by a single agent. Perspective shifts might be an interesting area of study. The scientific question then could be whether an agency's points of view is flowing through single organisations by single agents and whether these streams are reversible or not.

References

- BIERSCHENK, Bernhard (1984). *Steering mechanisms for knowability* (Kognitionsvetenskaplig forskning, No1). Lund: Lund University
- BIERSCHENK, Bernhard (1991). *The schema axiom as foundation of a theory for measurement and representation of consciousness* (Kognitionsvetenskaplig forskning, No. 38). Lund: Lund University. (ERIC Document Reproduction Service, No. ED 338 651, TM 017 364)
- BIERSCHENK, Bernhard. (1993). *The fundamentals of perspective text analysis* (Kognitionsvetenskaplig forskning, No. 45). Lund: Lund University, Department of Psychology.
- BIERSCHENK, Bernhard & BIERSCHENK, Inger (1976). *Computer-based content analysis*. Lund: Gleerup.
- BIERSCHENK, Bernhard, BIERSCHENK, Inger, & HELMERSSON, Helge (1996). 'Die Ökologie des Sprachraums.' In: Wilfried Bos, & Christian Tarnai (Eds.), *Computerunterstützte Inhaltsanalyse in den Empirischen Sozialwissenschaften* (pp. 11-30). Muenchen: Waxmann.
- BIERSCHENK, Inger (1977). *Datorbaserad innehållsanalys: Kodningsmanual* (Pedagogisk dokumentation, No. 52). Malmö: School of Education, Department of Educational and Psychological Research.
- BIERSCHENK, Inger (1984). *The schematism of natural language* (Kognitionsvetenskaplig forskning, No. 2). Lund: Lund University, Department of Psychology.
- BIERSCHENK, Inger (1989) *Language as carrier of consciousness* (Kognitionsvetenskaplig forskning, No. 30). Lund: Lund University, Department of Psychology.
- BIERSCHENK, Inger (1992). *The pendular movement of text building* (Kognitionsvetenskaplig forskning, No. 42). Lund: Lund University, Department of Psychology.
- BROADBENT, Donald, E. (1977). 'Levels, hierarchies and the locus of control.' *Quarterly Journal of Experimental Psychology*, 29: 181-201.
- CHOMSKY, Noam (1957). *Syntactic structures*. The Hague: Mouton.
- DIDERICHSEN, Poul (1962). *Elementær dansk grammatik*. Copenhagen: Gyldendal.
- ELSTRUP RASMUSSEN, Ole (1997). *Modelling Knowing Systems. Paper prepared for the Conference on Uncertainty, Knowledge and Skill, Limburg, November 6-8, 1997*. Copenhagen: University of Copenhagen, Psychological Laboratory and Copenhagen Competence Research Centre.
- FILLMORE, Charles, J. (1968). 'The case for case.' In Emmon Bach, & Richard, T. Harms (Eds.), *Universals in linguistic theory* (pp. 1-88). New York: Holt.

- HELMERSSON, Helge (1992). *Main principles for Perspective Text Analysis via the PC-system PERTEX* (Kognitionsvetenskaplig forskning, No. 41). Lund: Lund University, Department of Psychology.
- SINGH, J. D. (1979). 'Panini's theory of Karakas'. *Revue Roumaine de Linguistique*, 24/2: 123-148.